

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-356939  
(43)Date of publication of application : 26.12.2001

(51)Int.Cl.

G06F 11/34

(21)Application number : 2000-177205

(71)Applicant : TOKYO ELECTRIC POWER CO  
INC:THE

(22)Date of filing : 13.06.2000

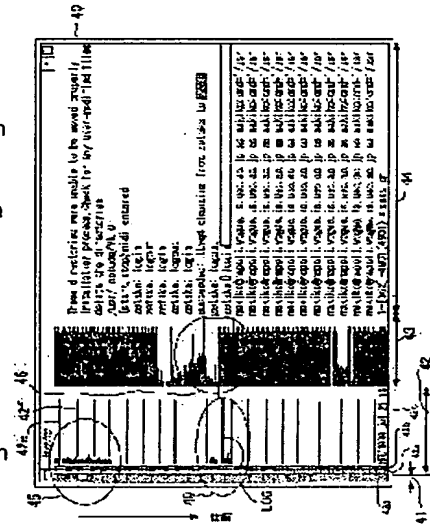
(72)Inventor : TAKADA TETSUJI  
KOIKE HIDEKI  
AMARI HARUO  
OKADA MIKIO

## (54) LOG INFORMATION ANALYZER, ITS METHOD AND RECORDING MEDIUM

### (57)Abstract:

**PROBLEM TO BE SOLVED:** To provide a log information analyzer, its method and a recording medium capable of reducing time to be required for an analyzing work of log information, integrating pieces of log information with various data forms and properties of uneven distribution on a file system, extracting log information to indicate an abnormality which is not the known one and enabling a system manager to exactly grasp noticeable log information among a vast quantity of log information by characters.

**SOLUTION:** Pieces of log information unevenly distributed in the log file are integrated on the basis of time information 17 by converting pieces of the log information recorded by various data forms into intermediate forms as multipurpose log formats. The log information to indicate the abnormality different from the known one is extracted by calculating appearance frequency 34 by word and appearance frequency 37 by phrase formed by linking two words. The time to be required for the analyzing work of the log information is reduced since the log information is combined with extracted feature information, defined as a graphic and offered to the system manager.



## LEGAL STATUS

[Date of request for examination] 09.11.2000

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(11)特許出願公開番号

特開2001-356939

(P2001-356939A)

(43)公開日 平成13年12月26日(2001.12.26)

(51) Int.Cl.<sup>7</sup>

識別記号

FI

テーマコード(参考)

G 0 6 F 11/34

G 0 6 F 11/34

C 5 B 0 4 2

審査請求 有 請求項の数17 O L (全 15 頁)

(21)出願番号 特願2000-177205(P2000-177205)

(22)出願日 平成12年6月13日(2000.6.13)

特許法第30条第1項適用申請有り 1999年12月15日 インターネットコンファレンス'99実行委員会発行の「インターネットコンファレンス'99論文集」に発表

(71)出願人 000003687

東京電力株式会社

東京都千代田区内幸町1丁目1番3号

(72)発明者 高田 哲司

東京都調布市飛田給2丁目46番12号 エミ  
ールマンション205

(72)発明者 小池 英樹

東京都杉並区荻窪5丁目29番17号 1104

(72)発明者 甘利 治雄

神奈川県横浜市鶴見区江ヶ崎町4番1号  
東京電力株式会社システム研究所内

(74) 代理人 100082175

弁理士 高田 守 (外2名)

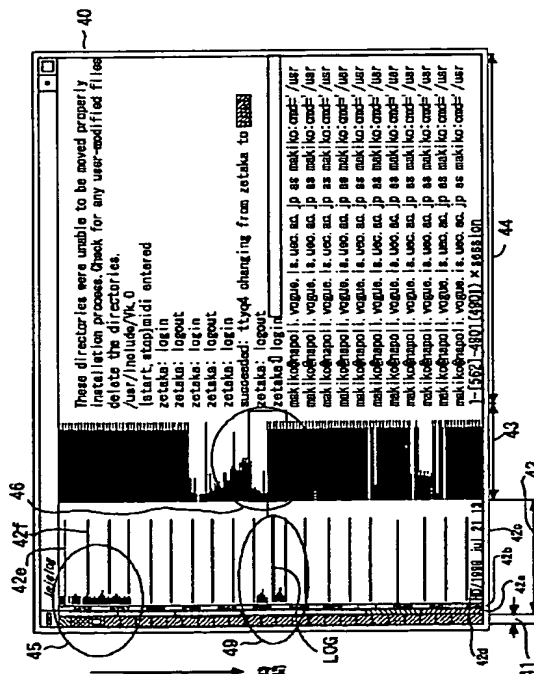
最終頁に続く

(54) 【発明の名称】 ログ情報解析装置、方法および記録媒体

(57) 【要約】

【課題】 ログ情報の解析作業に要する時間を減少させ、種々のデータ形式とファイル・システム上の偏在性とを有するログ情報を統合し、既知の異常ではない異常を示すログ情報を抽出することができ、システム管理者が文字による膨大な量のログ情報の中から注目すべきログ情報を迅速に把握することができるログ情報解析装置、方法および記録媒体を提供する。

【解決手段】 種々のデータ形式により記録されているログ情報を汎用ログフォーマットという中間形式に変換することにより、ログファイル中に偏在していたログ情報を時刻情報１７を基準として統合することができる。単語別出現頻度３４と二単語連結句別出現頻度３７とを求めることにより、既知の異常とは異なる異常を示すログ情報を抽出することができる。ログ情報と抽出された特徴情報とを組み合わせることで図形化してシステム管理者に提示することができるため、ログ情報の解析作業に要する時間を減少させることができる。



## 【特許請求の範囲】

【請求項1】 コンピュータの稼動状況を示すログ情報が記録されたログ情報記録部と、

前記ログ情報記録部に記録されたログ情報から該ログ情報が記録された時刻と所定の文字列とを抽出し、所定の時刻形式に変換された時刻と所定のフォーマットに変換された所定の文字列とを含む汎用ログを作成する汎用ログ作成手段と、

前記汎用ログ作成手段により作成された汎用ログを、前記所定の時刻形式に変換された時刻の順に記録する汎用ログ記録手段と、

前記汎用ログ記録手段により記録された汎用ログを有する汎用ログ記録部と、

前記汎用ログ記録部に記録された汎用ログに対応するログ情報の長さや該ログ情報中に現れる語の出現頻度とを含む特徴情報を前記ログ情報記録部から抽出する特徴情報抽出手段と、

前記特徴情報抽出手段により抽出されたログ情報を図形化して前記ログ情報と共に前記所定の時刻形式に変換された時刻の順に表示する表示手段とを備えたことを特徴とするログ情報解析装置。

【請求項2】 請求項1記載のログ情報解析装置において、前記特徴情報に含まれる語の出現頻度は、前記所定のフォーマットに変換された所定の文字列中に1単語が単独で出現する頻度と少なくとも2単語が連続して出現する頻度とを用いて求められることを特徴とするログ情報解析装置。

【請求項3】 請求項1記載のログ情報解析装置において、前記表示手段が表示するログ情報は、前記特徴情報に含まれる語の出現頻度が所定の値を有する語または指定された所定の語を少なくとも含む文字列が識別可能に表示されることを特徴とするログ情報解析装置。

【請求項4】 請求項1ないし3のいずれかに記載のログ情報解析装置において、前記表示手段は、前記所定のフォーマットに変換された所定の文字列が所定の種類に該当するログ情報を選択して表示することを特徴とするログ情報解析装置。

【請求項5】 請求項1ないし3のいずれかに記載のログ情報解析装置において、前記表示手段は、前記特徴情報に含まれる語の出現頻度が所定の範囲内にある語を含むログ情報を選択して表示することを特徴とするログ情報解析装置。

【請求項6】 請求項1ないし3のいずれかに記載のログ情報解析装置において、前記表示手段は、前記ログ情報の長さが所定の範囲内にあるログ情報を選択して表示することを特徴とするログ情報解析装置。

【請求項7】 請求項1ないし3のいずれかに記載のログ情報解析装置において、前記表示手段は、前記所定の時刻形式に変換された時刻が所定の範囲内にあるログ情報を選択して表示することを特徴とするログ情報解析装置。

置。

【請求項8】 請求項1ないし3のいずれかに記載のログ情報解析装置において、前記表示手段は、所定の文字列を有するログ情報を選択して表示することを特徴とするログ情報解析装置。

【請求項9】 請求項1ないし8のいずれかに記載のログ情報解析装置において、前記汎用ログ作成手段における前記所定の時刻形式は、所定の基準となる時刻から前記ログ情報が記録された時刻までの経過時間であることを特徴とするログ情報解析装置。

【請求項10】 請求項1ないし9のいずれかに記載のログ情報解析装置において、前記表示手段における前記ログ情報の図形化は、所定の時間あたりの該ログ情報の数のヒストグラム化であることを特徴とするログ情報解析装置。

【請求項11】 請求項1ないし9のいずれかに記載のログ情報解析装置において、前記表示手段における前記ログ情報の図形化は、該ログ情報に含まれる所定の文字列の長さに対応した線とすることを特徴とするログ情報解析装置。

【請求項12】 コンピュータの稼動状況を示すログ情報が記録されたログ情報記録部から該ログ情報が記録された時刻と所定の文字列とを抽出し、所定の時刻形式に変換された時刻と所定のフォーマットに変換された所定の文字列とを含む汎用ログを作成する汎用ログ作成ステップと、

前記汎用ログ作成ステップで作成された汎用ログを、前記所定の時刻形式に変換された時刻の順に汎用ログ記録部へ記録する汎用ログ記録ステップと、

前記汎用ログ記録部に記録された汎用ログに対応するログ情報の長さや該ログ情報中に現れる語の出現頻度とを含む特徴情報を前記ログ情報記録部に記録されたログ情報から抽出する特徴情報抽出ステップと、

前記特徴情報抽出ステップで抽出されたログ情報を図形化して前記ログ情報と共に前記所定の時刻形式に変換された時刻の順に表示する表示ステップとを備えたことを特徴とするログ情報解析方法。

【請求項13】 請求項12記載のログ情報解析方法において、前記特徴情報に含まれる語の出現頻度は、前記所定のフォーマットに変換された所定の文字列中に1単語が単独で出現する頻度と少なくとも2単語が連続して出現する頻度とを用いて求められることを特徴とするログ情報解析方法。

【請求項14】 請求項12記載のログ情報解析方法において、前記表示ステップで表示するログ情報は、前記特徴情報に含まれる語の出現頻度が所定の値を有する語または指定された所定の語を少なくとも含む文字列が識別可能に表示されることを特徴とするログ情報解析方法。

【請求項15】 コンピュータの稼動状況を示すログ情

報を解析するログ情報解析方法を実行するコンピュータが読み出し可能なプログラムを格納した記録媒体であって、

前記ログ情報が記録されたログ情報記録部から該ログ情報が記録された時刻と所定の文字列とを抽出し、所定の時刻形式に変換された時刻と所定のフォーマットに変換された所定の文字列とを含む汎用ログを作成する汎用ログ作成ステップと、

前記汎用ログ作成ステップで作成された汎用ログを、前記所定の時刻形式に変換された時刻の順に汎用ログ記録部へ記録する汎用ログ記録ステップと、

前記汎用ログ記録部に記録された汎用ログに対応するログ情報の長さや該ログ情報中に現れる語の出現頻度とを含む特徴情報を前記ログ情報記録ステップで記録されたログ情報から抽出する特徴情報抽出ステップと、前記特徴情報抽出ステップで抽出されたログ情報を図形化して前記ログ情報と共に前記所定の時刻形式に変換された時刻の順に表示する表示ステップとを備えたことを特徴とするログ情報解析方法を実行するコンピュータが読み出し可能なプログラムを格納した記録媒体。

【請求項16】 請求項15記載の記録媒体において、前記特徴情報に含まれる語の出現頻度は、前記変換された所定の文字列中に1単語が単独で出現する頻度と少なくとも2単語が連続して出現する頻度とを用いて求められることを特徴とする記録媒体。

【請求項17】 請求項15記載の記録媒体において、前記表示ステップで表示するログ情報は、前記特徴情報に含まれる語の出現頻度が所定の値を有する語または指定された所定の語を少なくとも含む文字列が識別可能に表示されることを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ログ情報解析装置、方法および記録媒体に関し、特にコンピュータ・システムの運用管理に用いられるログ情報解析装置、方法および記録媒体に関する。

【0002】

【従来の技術】コンピュータ・システムの運用管理、特に障害分析または異常現象の発見等を行うためには定期的なログ情報の調査、監視または解析等の作業が重要である。特に、近年ではコンピュータ・システムへの不正侵入を検知するためにログ情報の解析作業の重要性は極めて増大している。

【0003】しかし、従来のログ情報の解析作業は、システム管理者が既存のエディタ等を用いて膨大なログ情報を有するログファイルの閲覧を行い、独自のスクリプトを使用して特定のログを抽出していたため、極めて時間を要する作業となっていた。特に、ログ情報は種々のデータ形式を有している上にファイル・システム上に偏在しているという特徴があるため、ログ情報を解析する

前の調査時における作業負担が極めて重いという問題があった。

【0004】さらに、ログ情報は膨大な量を有しているため、システム管理者が注目すべきログ情報を抽出することは極めて困難である。既知の異常を示すログ情報の抽出は既存のキーワード検索命令を利用することにより可能ではあるが、ログ情報の中には既知の異常ではない異常を示すログ情報が含まれている場合もある。このような既知の異常ではない異常を示すログ情報を抽出することは極めて困難であるという問題があった。

【0005】ログ情報は文字による記録であるため、システム管理者が文字による膨大な量のログ情報の中から注目すべきログ情報を迅速に把握することは極めて困難であるという問題があった。

【0006】例えば、コンピュータ・システムへの不正侵入を検知する侵入検知システム（Intrusion Detection System：IDS）では、監視する不正侵入のパターン・データをシグネチャとして作成している。新しい不正侵入の方法が発見されるたびに新しい不正侵入のパターン・データをシグネチャに登録することにより不正侵入に対応している。この結果、既知の登録されているパターン・データとは異なる不正侵入が試みられた場合には、必ずしも有効に対応することができないという問題があった。膨大なログ情報の中から警報を抽出し、さらにその警報の中から誤報を切り捨てて注目すべき警報を迅速に把握することは極めて困難であるという問題があった。

【0007】

【発明が解決しようとする課題】上述のように、従来のログ情報の解析作業は極めて時間を要する作業であり、ログ情報の種々のデータ形式とファイル・システム上の偏在性という特徴のため作業負担が極めて重いという問題があった。さらに、膨大なログ情報の中から既知の異常ではない異常を示すログ情報を抽出することは極めて困難であるという問題があった。システム管理者が文字による膨大な量のログ情報の中から注目すべきログ情報を迅速に把握することは極めて困難であるという問題があった。

【0008】そこで、本発明の目的は、上記問題を解決するためになされたものであり、ログ情報の解析作業に要する時間を減少させ、種々のデータ形式とファイル・システム上の偏在性とを有するログ情報を統合し、既知の異常ではない異常を示すログ情報を抽出することができ、システム管理者が文字による膨大な量のログ情報の中から注目すべきログ情報を迅速に把握することができるログ情報解析装置、方法および記録媒体を提供することにある。

【0009】

【課題を解決するための手段】請求項1に記載された発明のログ情報解析装置は、コンピュータの稼動状況を示

すログ情報が記録されたログ情報記録部と、前記ログ情報記録部に記録されたログ情報から該ログ情報が記録された時刻と所定の文字列とを抽出し、所定の時刻形式に変換された時刻と所定のフォーマットに変換された所定の文字列とを含む汎用ログを作成する汎用ログ作成手段と、前記汎用ログ作成手段により作成された汎用ログを、前記所定の時刻形式に変換された時刻の順に記録する汎用ログ記録手段と、前記汎用ログ記録手段により記録された汎用ログを有する汎用ログ記録部と、前記汎用ログ記録部に記録された汎用ログに対応するログ情報の長さとして該ログ情報中に現れる語の出現頻度とを含む特徴情報を前記ログ情報記録部から抽出する特徴情報抽出手段と、前記特徴情報抽出手段により抽出されたログ情報を図形化して前記ログ情報と共に前記所定の時刻形式に変換された時刻の順に表示する表示手段を備えたものである。

【0010】請求項2に記載された発明のログ情報解析装置は、請求項1において、前記特徴情報に含まれる語の出現頻度は、前記所定のフォーマットに変換された所定の文字列中に1単語が単独で出現する頻度と少なくとも2単語が連続して出現する頻度とを用いて求めることができる。

【0011】請求項3に記載された発明のログ情報解析装置は、請求項1において、前記表示手段が表示するログ情報は、前記特徴情報に含まれる語の出現頻度が所定の値を有する語または指定された所定の語を少なくとも含む文字列が識別可能に表示されることができる。

【0012】請求項4に記載された発明のログ情報解析装置は、請求項1ないし3のいずれかにおいて、前記表示手段は、前記所定のフォーマットに変換された所定の文字列が所定の種類に該当するログ情報を選択して表示することができる。

【0013】請求項5に記載された発明のログ情報解析装置は、請求項1ないし3のいずれかにおいて、前記表示手段は、前記特徴情報に含まれる語の出現頻度が所定の範囲内にある語を含むログ情報を選択して表示することができる。

【0014】請求項6に記載された発明のログ情報解析装置は、請求項1ないし3のいずれかにおいて、前記表示手段は、前記ログ情報の長さが所定の範囲内にあるログ情報を選択して表示することができる。

【0015】請求項7に記載された発明のログ情報解析装置は、請求項1ないし3のいずれかにおいて、前記表示手段は、前記所定の時刻形式に変換された時刻が所定の範囲内にあるログ情報を選択して表示することができる。

【0016】請求項8に記載された発明のログ情報解析装置は、請求項1ないし3のいずれかにおいて、前記表示手段は、所定の文字列を有するログ情報を選択して表示することができる。

【0017】請求項9に記載された発明のログ情報解析装置は、請求項1ないし8のいずれかにおいて、前記汎用ログ作成手段における前記所定の時刻形式は、所定の基準となる時刻から前記ログ情報が記録された時刻までの経過時間とすることができる。

【0018】請求項10に記載された発明のログ情報解析装置は、請求項1ないし9のいずれかにおいて、前記表示手段における前記ログ情報の図形化は、所定の時間あたりの該ログ情報の数のヒストグラム化とすることができる。

【0019】請求項11に記載された発明のログ情報解析装置は、請求項1ないし9のいずれかにおいて、前記表示手段における前記ログ情報の図形化は、該ログ情報に含まれる所定の文字列の長さに対応した線とすることができる。

【0020】請求項12に記載された発明のログ情報解析方法は、コンピュータの稼動状況を示すログ情報が記録されたログ情報記録部から該ログ情報が記録された時刻と所定の文字列とを抽出し、所定の時刻形式に変換された時刻と所定のフォーマットに変換された所定の文字列とを含む汎用ログを作成する汎用ログ作成ステップと、前記汎用ログ作成ステップで作成された汎用ログを、前記所定の時刻形式に変換された時刻の順に汎用ログ記録部へ記録する汎用ログ記録ステップと、前記汎用ログ記録部に記録された汎用ログに対応するログ情報の長さとして該ログ情報中に現れる語の出現頻度とを含む特徴情報を前記ログ情報記録部に記録されたログ情報から抽出する特徴情報抽出ステップと、前記特徴情報抽出ステップで抽出されたログ情報を図形化して前記ログ情報と共に前記所定の時刻形式に変換された時刻の順に表示する表示ステップとを備えたものである。

【0021】請求項13に記載された発明のログ情報解析方法は、請求項12において、前記特徴情報に含まれる語の出現頻度は、前記所定のフォーマットに変換された所定の文字列中に1単語が単独で出現する頻度と少なくとも2単語が連続して出現する頻度とを用いて求めることができる。

【0022】請求項14に記載された発明のログ情報解析方法は、請求項12において、前記表示ステップで表示するログ情報は、前記特徴情報に含まれる語の出現頻度が所定の値を有する語または指定された所定の語を少なくとも含む文字列が識別可能に表示されることができる。

【0023】請求項15に記載された発明の記録媒体は、コンピュータの稼動状況を示すログ情報を解析するログ情報解析方法を実行するコンピュータが読み出し可能なプログラムを格納した記録媒体であって、前記ログ情報が記録されたログ情報記録部から該ログ情報が記録された時刻と所定の文字列とを抽出し、所定の時刻形式に変換された時刻と所定のフォーマットに変換された所

定の文字列とを含む汎用ログを作成する汎用ログ作成ステップと、前記汎用ログ作成ステップで作成された汎用ログを、前記所定の時刻形式に変換された時刻の順に汎用ログ記録部へ記録する汎用ログ記録ステップと、前記汎用ログ記録部に記録された汎用ログに対応するログ情報の長さや該ログ情報中に現れる語の出現頻度とを含む特徴情報を前記ログ情報記録ステップで記録されたログ情報から抽出する特徴情報抽出ステップと、前記特徴情報抽出ステップで抽出されたログ情報を図形化して前記ログ情報と共に前記所定の時刻形式に変換された時刻の順に表示する表示ステップとを備えたログ情報解析方法を実行するコンピュータが読み出し可能なプログラムを格納した記録媒体である。

【0024】請求項16に記載された発明の記録媒体は、請求項15において、前記特徴情報に含まれる語の出現頻度は、前記変換された所定の文字列中に1単語が単独で出現する頻度と少なくとも2単語が連続して出現する頻度とを用いて求めることができる。

【0025】請求項17に記載された発明の記録媒体は、請求項15において、前記表示ステップで表示するログ情報は、前記特徴情報に含まれる語の出現頻度が所定の値を有する語または指定された所定の語を少なくとも含む文字列が識別可能に表示されることができる。

【0026】

【発明の実施の形態】以下、図1を用いてまず本発明の各実施の形態に共通するログ情報解析装置の概要を示し、次に図面を参照して各実施の形態を詳細に説明する。

【0027】図1は、本発明のログ情報解析装置に用いられるコンピュータの内部回路ブロックを示す。図1において、符号15は本発明のログ情報解析装置の内部回路ブロック、1は本発明のログ情報解析方法等を実行する処理装置CPU (Central Processing Unit)、2は内部回路ブロック15の初期化等その他の処理に必要なデータが格納された読み出し専用記憶装置ROM (Read Only Memory)、3はCPU1が実行するコンピュータ・プログラムまたはデータが格納された読み書き可能な記憶装置RAM (Random Access Memory)、4はCPU1に処理結果に基づいて画像を表示するディスプレイ等の画像表示部、6はログ情報解析方法等を実行するためのコンピュータ・プログラムまたはデータ等を記録したコンピュータ読み取り可能なCD-ROM (Compact Disc - Read Only memory) 等の脱着可能な記録媒体をセットする記録媒体部、8はログ情報が記録されたログファイル (ログ情報記録手段)、9はログファイル8から所望のデータを抽出して所望の変換が行われた汎用ログが記録された汎用ログファイル (汎用ログ情報記録手段)、7はログファイル8と汎用ログファイル9とが記録されたハード・ディスク等の記録装置、5は記録媒体部6または記録装置7等と接続され入出力の制御を行う

入出力制御部、11はログ情報解析装置へ入力操作を行うマウス、キーボード等の入力操作部、10は入力操作部11と接続され入力制御等を行う入力制御部、12はネットワーク (不図示) を介して外部のコンピュータ等と行なう通信を制御する通信制御部、14は上述のCPU1、ROM2、RAM3、画像表示部4、入出力制御部5、入力制御部10および通信制御部12を接続するバスである。

【0028】本発明のログ情報解析方法等を実行するコンピュータ・プログラムおよびデータは記録媒体部6にセットされたCD-ROM等の記録媒体に記録しておくことができる。CD-ROMまたはメモリ・カード等の記録媒体に記録された上記コンピュータ・プログラムおよびデータは、入出力制御部5を介してバス14を通りRAM3へロードされる。CPU1はRAM3内にロードされた上記コンピュータ・プログラムを実行することにより、入力操作部11から入力制御部10を介して所望の入力が行なわれ、画像表示部4にログ情報解析方法を実行中の画像を表示させることができる。

【0029】実施の形態1. 図2は、本発明の実施の形態1におけるログ情報解析方法のフローチャートを示す。以下、フローチャートの各ステップを図面を参照しながら説明する。図2に示されるように、まずログファイル8からログ情報を入力する (ステップS102)。このログ情報はコンピュータの稼動状況を示す情報である。以下では説明の都合上、不正アクセスを検知するIDS等におけるログ情報を用いて説明しているが、本発明のログ情報解析方法等が対象とするログ情報は一般的なログ情報であって、IDS等におけるログ情報に限定されるものではない。

【0030】次に、入力したログ情報を後述の汎用ログフォーマットに変換する (ステップS104、汎用ログ作成手段)。ログ情報は、ログ情報がログファイル8に記録された時刻と、そのログ情報の内容、例えばどこから接続されたか等の内容を有するデータ (所定の文字列) とを有している。このログ情報から時刻と他の所望のデータを抽出し、汎用ログフォーマットにしたがって汎用ログを作成する。

【0031】図3は、本発明の実施の形態1における汎用ログフォーマットを示す。図3において、符号16は汎用ログフォーマットの全体を示し、17は所定の時刻形式に変換された時刻情報、18および19は各々空白および指定の区切り文字を含まない任意の文字列であるタグ情報1およびタグ情報2、20は指定の区切り文字を含まない任意の文字列であるメッセージ情報である。

【0032】図4は本発明の実施の形態1におけるログ情報から汎用ログへの変換方法のフローチャートを示し、図5 (A)、(B) は本発明の実施の形態1におけるログ情報から汎用ログへの変換例を示す。図5で図3と同じ符号を付した箇所は同じ部分を示すため説明は省

略する。

【0033】図4に示されるように、まず時刻情報17とする情報をログ情報から1つ決定する(ステップS202)。例えば図5(A)に示されるように、時刻情報17とする情報をログ情報ACCT29Aの時刻23(03:17:15)と決定する。

【0034】次に、決定された時刻23を所定の基準となる時刻、例えば1970年1月1日9:00からの経過秒数(所定の時刻形式)へ変換する(ステップS204)。ここで所定の基準となる時刻は任意に設定することができる。この結果、例えば図5(A)に示されるように、時刻情報17の値(変換された時刻)は、「952625865(秒)」と設定される。

【0035】ログ情報から、後述のフィルタリングまたは単語の出現頻度を抽出する際の基礎情報となる情報を決定し、所定の文字列へ変換することによりタグ情報1(18)を作成する(ステップS206)。タグ情報2(19)も同様に作成する(ステップS208)。例えば図5(A)に示されるように、基礎となる情報をログ情報ACCT29Aの文字列「ypwhich」21と決定し、この文字列21をそのままタグ情報2(19)として作成する。タグ情報1(18)にはログ情報ACCT28に含まれていない情報、例えばログ情報解析装置の実行環境から得られる補完的な情報(例、「foo.co.jp」)を用いることもできる。

【0036】最後に、ログ情報から解析に必要となる情報を決定し、所定の文字列へ変換することによりメッセージ情報20を作成する(ステップS210)。例えば図5(A)に示されるように、解析に必要となる情報をログ情報ACCT29Aの文字列「root?」22と文字列「0.05 0.01 540.00」24とに決定し、これらの文字列22と24とを組み合わせた新たな文字列「root? 0.05 0.01 540.00」をメッセージ情報20として作成する。

【0037】図5(B)は他の変換例を示す。図5(B)に示されるように、時刻情報17とする情報をログ情報SYSLOG29Bの時刻25(Oct25 10:41:25)と決定する。決定された時刻25を所定の基準となる時刻からの経過秒数へ変換し、例えば図5(B)に示されるように、時刻情報17の値を「933912865(秒)」と設定する。

【0038】タグ情報1(18)は、図5(B)に示されるように基礎となる情報をログ情報SYSLOG29Bの文字列「foo.co.jp」26と決定し、この文字列26をそのままタグ情報1(18)として作成する。タグ情報2(19)は、図5(B)に示されるように基礎となる情報をログ情報SYSLOG29Bの文字列「in.telnetd」27と決定し、この文字列27をそのままタグ情報2(19)として作成する。

【0039】メッセージ情報20は、図5(B)に示さ

れるように、解析に必要となる情報をログ情報SYSLOG29Bの文字列「connect from crack.com」28と決定し、この文字列28をそのままメッセージ情報20として作成する。

【0040】上述のように、ログ情報を汎用ログフォーマットという中間形式に変換することにより、ファイル・システム中に偏在していたログ情報を時刻情報17を基準として統合することができる。

【0041】図2のフローチャートに戻り、ステップS104に続いて、特徴情報を抽出する(特徴情報抽出手段、ステップS106)。特徴情報は、汎用ログフォーマットに変換された時刻情報17、タグ情報1(18)、タグ情報2(19)またはメッセージ情報(20)等またはログ情報から抽出する情報である。後述するように、これらの特徴情報をログ情報と共に視覚化して画像表示部4に表示することにより、システム管理者は的確かつ迅速にログ情報を把握することができる。特徴情報には、単位時間当たりのログ情報の数、タグ情報1(18)等の種類別のログ情報の数、メッセージ情報20の文字列の長さまたはメッセージ情報20に含まれる単語の出現頻度等がある。上述の特徴情報は例示的なものであり、これらの例に限定されるものではない。

【0042】図6は、本発明の実施の形態1におけるメッセージ情報20に含まれる単語の出現頻度を求める方法を説明する。図6に示されるように、ログメッセージ30に対してテキストマイニングによる情報抽出33を行い、1単語が単独で出現する単語別出現頻度34と2単語が連続して出現する二単語連結句別出現頻度37とを得ることができる。得られた単語別頻度34と二単語連結句別出現頻度37とは、各々図6に示されるように、出現数と単語等との一覧により示されている。

【0043】例えば、単語別出現頻度34に示されるように、単語「from」はログメッセージ30中に6回単独で出現しており、一方、単語「kyoto」35はログメッセージ30中に1回単独で出現し、単語「sendai」36はログメッセージ30中に1回単独で出現している。二単語連結句別出現頻度37に示されるように、二単語「connect from」はログメッセージ30中に4回出現しており、一方、二単語「from kyoto」38はメッセージ30中に1回単独で出現し、単語「from sendai」39はログメッセージ30中に1回単独で出現している。この結果、単語別出現頻度34において最も出現頻度の低い単語「kyoto」35および単語「sendai」36と、二単語連結句別出現頻度37において最も出現頻度の低い二単語「connect from」および二単語「from kyoto」38とを含むログメッセージ31と32とが調査を要するログメッセージであることを推測することができる。

【0044】一般的なログメッセージは数が多く記録さ

れているため、記録されている数が少ない、すなわち出現頻度の低いログメッセージは不正侵入等を示す可能性等が高いログメッセージであると考えることができる。本発明の実施の形態1では上述のように単語別出現頻度34と二単語連結句別出現頻度37とを求めることにより、不正侵入等の、既知の異常とは異なる異常を示すログ情報を抽出することができ、注目すべきログ情報をシステム管理者に提示することができる。

【0045】図2のフローチャートに戻り、ステップS106に続いて、ログ情報を視覚化して表示する（表示手段、ステップS108）。具体的には、単位時間当たりのログ情報の数をヒストグラムとして表示し、ログ情報の文字列の長さに対応した線を用いてログ情報をアウトライン的に表示し（アウトライン表示）、並列的にログ情報の文字の情報の部分を表示する。その他、タグ情報1（18）等別のログ情報の数を表示することもできる。

【0046】図7は、本発明の実施の形態1におけるログ情報の視覚化表示を示す。図7において、符号40はログ情報の視覚化表示の全体を示し、ログ情報は縦軸方向の矢印（時間軸）で示されるように記録された時刻の順に上から下へ表示されている。符号41は注目しているログのタグ情報1（18）等の頻度を示すタグ格子領域またはタグ種別表示領域を示す。このタグ格子領域41は図7では明示されていないが、最上部が青色で、以下黄色になり、最下部になるほど赤色で表示されている。この色分けは、後述する他の領域42等における各種の表示において、出現頻度が多いものほど青色で表示され、出現頻度が中程度であるものは黄色で表示され、出現頻度が低いものほど注意を引くために赤色で表示することを示すための領域である。上述の色分けは例示的なものであって、出現頻度の区別ができる色分けであれば任意の色分けを用いることができる。

【0047】図7において、符号42はログ情報の数の時間分布を表示する時間別頻度領域または時刻情報表示領域である。時間別頻度領域42は、時間軸方向に曜日に分かれた週格子42aと時間に分かれた日格子42bとを有しており、横軸方向にログ情報の数が示された単位時間別のヒストグラムを表示するヒストグラム領域42cを有している。ヒストグラム領域42cは時間軸順に上から下へ等間隔で横線42e、42f等が引かれている。これらの横線42eと42fとの間は12時間分の時間間隔を示している。符号45の円内に示されるヒストグラムは所定の期間にわたってほぼ一定の数のログ情報が記録されたことを示している。週格子42aと日格子42bとを用いることにより、領域42cに表示されたログ情報のヒストグラムの曜日毎、時間毎の周期的な側面の把握をさらに容易に行うことができる。符号49の円内に示されるヒストグラムは極めて少ない数のログ情報の記録がなされている間に、符号LOGで示され

る極めて膨大な数のログが記録されたことを示している。このログ（LOG）を以下では注目ログと呼ぶ。符号42dは注目ログLOGが記録された時刻を示している。

【0048】図7において、符号43はログ情報の文字列の長さに対応した線でアウトライン的にログ情報を表示するアウトライン表示領域である。アウトライン表示領域については後述する。符号44はアウトライン表示領域43における注目ログLOGを含む注目領域46内の実際のログ情報を拡大して表示するログメッセージ表示領域である。注目ログLOGを時刻情報表示領域42内でマウス等の入力操作部11によりクリック等して選択すると、注目ログLOGを中心として時間軸上の前後の概要をアウトライン表示領域43に表示することができる。アウトライン表示領域43中の注目領域46の部分が拡大されてログメッセージ表示領域44に対応する実際のログ情報を表示することができる。

【0049】図8（A）、（B）は本発明の実施の形態1におけるアウトライン表示領域を説明する。図8で図7と同じ符号を付した個所は同じ部分を示すため説明は省略する。図8（A）において、符号50はログメッセージ表示領域44におけるログメッセージを一部取り出したものであり、50aないし50dは各ログメッセージを示す。符号51はログメッセージ50中の各ログメッセージ50a等を文字列の長さに対応した長さを有する線として表示したアウトライン表示である。アウトライン表示51aないし51dの各線はログメッセージ50aないし50dに対応している。このアウトライン表示51a等の長さに加えて、タグ種別表示領域41に示されるログのタグ情報1（18）等の頻度に応じた色分けをさらに施すことができる。図8（A）では色分けの代わりにハッチング等の種類で頻度を表示している。

【0050】図8（B）において、符号52はログ情報の視覚化表示の全体40からアウトライン表示領域43とログメッセージ表示領域44とを取り出した表示を示す。アウトライン表示領域43における注目ログLOGを含む注目領域46内の実際のログ情報が拡大されて、ログメッセージ表示領域44内の領域54内に表示されている。逆に、ログメッセージ表示領域44内の領域54内に表示されている文字情報が、注目領域46にログ情報の長さを基に抽象化された図形となって表示されていると見ることもできる。

【0051】図9は、本発明の実施の形態1におけるログ情報の視覚化表示の他の例を示す。図9で図7と同じ符号を付した個所は同じ部分を示すため説明は省略する。図9において、符号71、72および73はメッセージ情報20に含まれる単語の出現頻度（図6参照）に応じて色分けして表示されたログ情報である。図9では色分け表示の代わりに斜線をかぶせて示されているが、タグ種別表示領域41に示された頻度に応じた色分けを



施して表示することができる。上述のように、ログ情報と抽出された特徴情報とを組み合わせる図形化してシステム管理者に提示することができるため、ログ情報の解析作業に要する時間を減少させることができ、システム管理者は文字による膨大な量のログ情報の中から注目すべきログ情報を迅速に把握することができる。

【0052】以上より、実施の形態1によれば、種々のデータ形式により記録されているログ情報を汎用ログフォーマットという中間形式に変換することにより、ファイル・システム中に偏在していたログ情報を時刻情報17を基準として統合することができる。単語別出現頻度34と二単語連結句別出現頻度37とを求めることにより、不正侵入等の、既知の異常とは異なる異常を示すログ情報を抽出することができ、注目すべきログ情報をシステム管理者に提示することができる。ログ情報と抽出された特徴情報とを組み合わせる図形化してシステム管理者に提示することができるため、ログ情報の解析作業に要する時間を減少させることができ、システム管理者は文字による膨大な量のログ情報の中から注目すべきログ情報を迅速に把握することができる。

【0053】実施の形態2。図10は、本発明の実施の形態2におけるログ情報の視覚化表示の別の例を示す。図10で図9と同じ符号を付した箇所は同じものを示すため説明は省略する。図10において、符号70はログメッセージ表示領域44中のログ情報の視覚化表示、71ないし76はシステム管理者等から指定されたパターンマッチング用の文字列（指定された所定の語）を用いて抽出されたキーワード、77ないし82は上述の単語の出現頻度（図6参照）の解析により所定の範囲内の出現頻度を有するものとして指定されたキーワードである。図10では明示されていないが、キーワード71ないし76は赤い背景色を有するように表示することができる。キーワード77ないし82は青い背景色を有するように表示することができる。図10では色分けの変わりに、キーワード71ないし76は左斜線をかおせるように表示され、キーワード77ないし82は右斜線をかおせるように表示されている。赤い背景と青い背景とは例示的な色分けであり、他の任意の色分け、反転表示等を用いることもできる。

【0054】以上より、実施の形態2によれば、システム管理者等から指定されたパターンマッチング用の文字列を用いて抽出されたキーワードと、上述の単語の出現頻度が所定の範囲内であるものとして指定されたキーワードとを色分けして表示することができる。このため、システム管理者は文字による膨大な量のログ情報の中から注目すべきログ情報を迅速に把握することができる。

【0055】実施の形態3。図11は、本発明の実施の形態3におけるタグ情報による情報検索機能（フィルタリング）の例を示す。図11で図9と同じ符号を付した

箇所は同じものを示すため説明は省略する。図11において、符号85はフィルタリング処理を行う前の画面、86は指定されたタグ情報1（18）等の種類でフィルタリングを行った結果の画面、87は指定されたタグ情報1（18）の出現頻度でフィルタリングを行った結果の画面である。画面86に示されるように、タグ情報1（18）の種類の指定はタグ種別表示領域41中の特定の領域88をマウス等の入力操作部11でクリック等して選択することにより指定することができる。図11は、タグ情報1（18）として「inetd」が指定された例を示している。画面86に示されるように、タグ情報1（18）の出現頻度の指定はタグ種別表示領域41中の特定の領域89をマウス等の入力操作部11でドラッグ等して選択することにより指定することができる。

【0056】以上より、実施の形態3によれば、マウス等の入力操作部11でタグ情報1（18）等の種類またはタグ情報1（18）の出現頻度を選択することにより、容易にログ情報のフィルタリングを行うことができる。

【0057】実施の形態4。図12は、本発明の実施の形態4におけるメッセージ長によるフィルタリングの例を示す。図12で図9と同じ符号を付した箇所は同じものを示すため説明は省略する。図12において、符号90はフィルタリング処理を行う前の画面、92、94は指定された閾値のメッセージ長でフィルタリングを行った結果の画面である。画面92に示されるように、メッセージ長の閾値はアウトライン表示領域43内でメッセージ長閾値線91を横軸上でマウス等の入力操作部11を利用して左右に移動させることにより指定することができる。画面92では、注目領域93中でメッセージ長閾値線91より長いメッセージ長が抽出された例を示している。画面94では、注目領域96中でメッセージ長閾値線95より短いメッセージ長が抽出された例を示している。

【0058】以上より、実施の形態4によれば、アウトライン表示領域43内でメッセージ長閾値線91を横軸上でマウス等の入力操作部11を利用して左右に移動させることにより、容易に所望のメッセージ長より長いまたは短いメッセージ長のログ情報のフィルタリングを行うことができる。

【0059】実施の形態5。図13は、本発明の実施の形態5における時間帯によるフィルタリングの例を示す。図13で図9と同じ符号を付した箇所は同じものを示すため説明は省略する。図13において、符号100はフィルタリング処理を行う前の画面、102は指定された時間帯でフィルタリングを行った結果の画面である。画面100は、注目ログLOGを中心とした前後の時間帯をアウトライン表示領域43内の領域101に示している。この状態で、例えば時刻情報表示領域42内

の注目ログLOGより時間軸上で下の位置をマウス等の入力操作部11でクリック等することにより、画面102に示されるように注目ログLOGより後の時間帯（アウトライン領域43内の領域105）を表示することができる。時間軸上で上の位置をクリック等した場合は、注目ログLOGより前の時間帯を表示することができる。

【0060】以上より、実施の形態5によれば、時刻情報表示領域42内の注目ログLOGより時間軸上で上または下の位置をマウス等の入力操作部11でクリック等することにより、注目ログLOGより前または後の時間帯を表示することができる。

【0061】実施の形態6、図14は、本発明の実施の形態6における単語によるフィルタリングの例を示す。図14で図9と同じ符号を付した個所は同じものを示すため説明は省略する。図14において、符号110はフィルタリング処理を行う前の画面、112は指定された単語でフィルタリングを行った結果の画面である。画面100で例えば単語「zetaka」をマウス等の入力操作部11でドラッグ等して選択することにより、画面112に示されるように単語「zetaka」を含むログ情報のみを表示することができる。選択される単語は1個のみではなく複数の単語を指定することもできる。この場合、複数の単語はすべて含むか（論理積）または少なくとも1個を含むか（論理和）という指定をすることもできる。

【0062】以上より、実施の形態6によれば、所望の単語をマウス等の入力操作部11でドラッグ等して選択することにより、選択された単語を含むログ情報のみを表示することができる。

【0063】実施の形態7、上述した各実施の形態の機能を実現するコンピュータ・プログラムを記録した記録媒体を本発明のログ情報解析装置に供給し、そのログ情報解析装置のCPU1が記録媒体部6等にセットされた着脱可能な記録媒体に格納されたコンピュータ・プログラムを読み取り実行することによっても、本発明の目的が達成されることは言うまでもない。この場合、上述の記録媒体から読み取られたコンピュータ・プログラム自体が本発明のログ情報解析装置の新規な機能を実現することになり、そのコンピュータ・プログラムを記録した記録媒体は本発明を構成することになる。コンピュータ・プログラムを記録した記録媒体としては、例えば、CD-ROM、MO、メモ리카ード、光ディスク、フロッピー（登録商標）ディスク、ハードディスク、ROM等を用いることができる。

【0064】以上より、実施の形態7によれば、上述した各実施の形態の機能を実現するコンピュータ・プログラムを記録した記録媒体を本発明のログ情報解析装置に供給し、そのログ情報解析装置のCPU1が記録媒体に格納されたコンピュータ・プログラムを読み取り実行す

ることによっても、本発明の目的を達成することができる。

【0065】

【発明の効果】以上説明したように、本発明のログ情報解析装置、方法および記録媒体によれば、ログ情報を汎用ログフォーマットに変換し、単語別出現頻度34と二単語連結句別出現頻度37とを求め、ログ情報と抽出された特徴情報とを組み合わせる図形化してシステム管理者に提示することにより、ログ情報の解析作業に要する時間を減少させ、種々のデータ形式とファイル・システム上の偏在性とを有するログ情報を統合し、既知の異常ではない異常を示すログ情報を抽出することができ、システム管理者が文字による膨大な量のログ情報の中から注目すべきログ情報を迅速に把握することができるログ情報解析装置、方法および記録媒体を提供することができる。

【図面の簡単な説明】

【図1】 本発明のログ情報解析装置に用いられるコンピュータの内部回路ブロックを示す図である。

【図2】 本発明の実施の形態1におけるログ情報解析方法を示すフローチャートである。

【図3】 本発明の実施の形態1における汎用ログフォーマットを示す図である。

【図4】 本発明の実施の形態1におけるログ情報から汎用ログへの変換方法を示すフローチャートである。

【図5】 本発明の実施の形態1におけるログ情報から汎用ログへの変換例を示す図である。

【図6】 本発明の実施の形態1におけるメッセージ情報20に含まれる単語の出現頻度を求める方法を説明する図である。

【図7】 本発明の実施の形態1におけるログ情報の視覚化表示を示す図である。

【図8】 本発明の実施の形態1におけるアウトライン表示領域を説明する図である。

【図9】 本発明の実施の形態1におけるログ情報の視覚化表示の他の例を示す図である。

【図10】 本発明の実施の形態2におけるログ情報の視覚化表示の別の例を示す図である。

【図11】 本発明の実施の形態3におけるタグ情報によるフィルタリングの例を示す図である。

【図12】 本発明の実施の形態4におけるメッセージ長によるフィルタリングの例を示す図である。

【図13】 本発明の実施の形態5における時間帯によるフィルタリングの例を示す図である。

【図14】 本発明の実施の形態6における単語によるフィルタリングの例を示す図である。

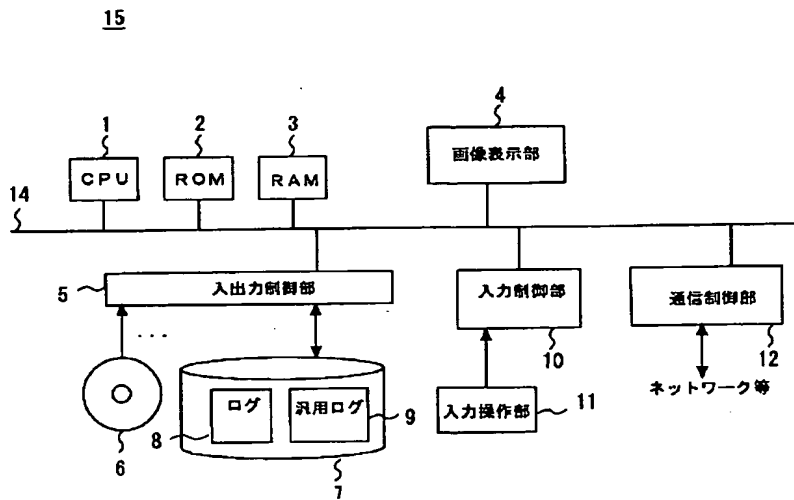
【符号の説明】

1 CPU、2 ROM、3 RAM、4 画像表示部、5 入出力制御部、6 記録媒体部、7 記録装置、8 ログ情報、9 汎用ログ情報、

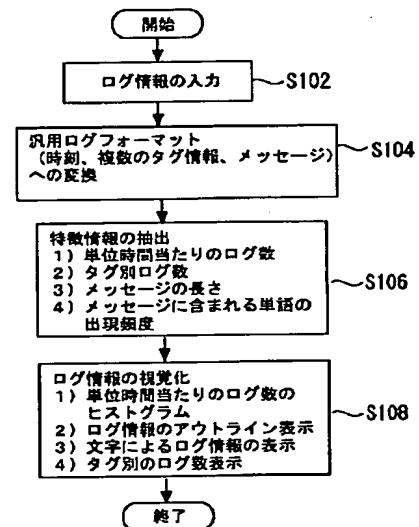
10 入力制御部、11 入力操作部、12 通信制御部、13 ネットワーク、14 バス、16 汎用ログフォーマットの全体、17 時刻情報、18 タグ情報1、19 タグ情報2、20 メッセージ情報、21、22、23、24、25、26、27 文字列、29A、29B ログ情報、30 ログメッセージ、31、32 ログメッセージ、33 テキストマイニングによる情報抽出、34 単語別出現頻度、35、36 単語、37 二単語連結句別出現頻度、38、39 二単語、41 タグ格子領域またはタグ種別表示領域、42 時間別頻度領域または時刻情報表示領域、42a 週格子、42b 日格子、42c ヒストグラム領域、42d 注目ログLOGが記録された時刻、42e、42f 横線、43 アウトライン表示領域、44 ログメッセージ表示領域、45、49 円、46、93、96 注目領域、50 ログメッセージの一部、50a、50b、50c、50d ログメッセージ、51、51a、51

b アウトライン表示、52 アウトライン表示領域43とログメッセージ表示領域44とを取り出した表示、54 領域、71、72、73 ログ情報、70 ログ情報の視覚化表示、71、72、73、74、75、76 パターンマッチング用の文字列を用いて抽出されたキーワード、77、78、79、80、81、82 指定されたキーワード、85、90、100、110 フィルタリング処理を行う前の画面、86 指定されたタグ情報1(18)等の種類でフィルタリングを行った結果の画面、87 指定されたタグ情報1(18)の出現頻度でフィルタリングを行った結果の画面、91、95 メッセージ長閾値線、92、94 指定された閾値のメッセージ長でフィルタリングを行った結果の画面、102 指定された時間帯でフィルタリングを行った結果の画面、101、105 領域、112 指定された単語でフィルタリングを行った結果の画面。

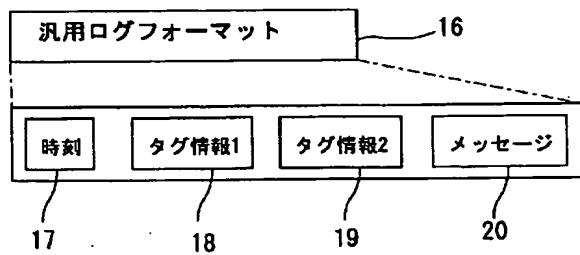
【図1】



【図2】

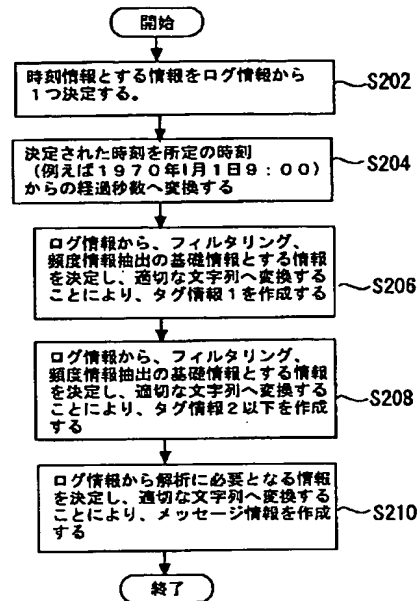


【図3】

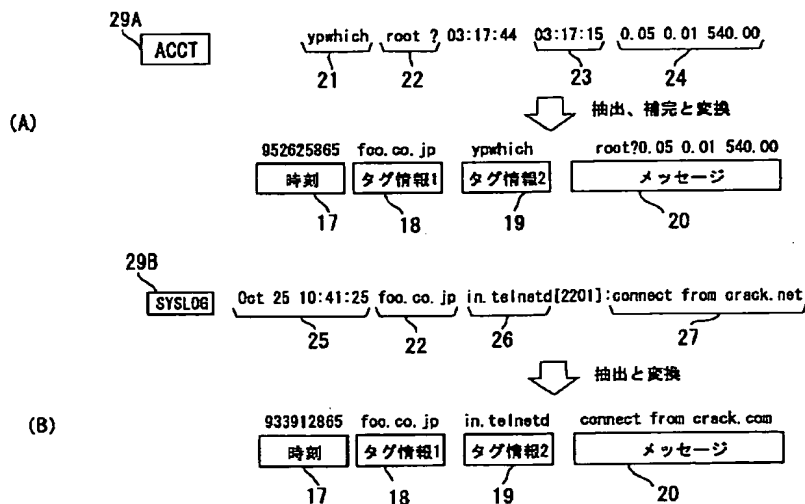


17: 1970年1月1日9:00からの経過秒数  
 18: 空白および指定の区切り文字以外の任意の文字列  
 19: 空白および指定の区切り文字以外の任意の文字列  
 20: 指定の区切り文字以外の任意の文字列

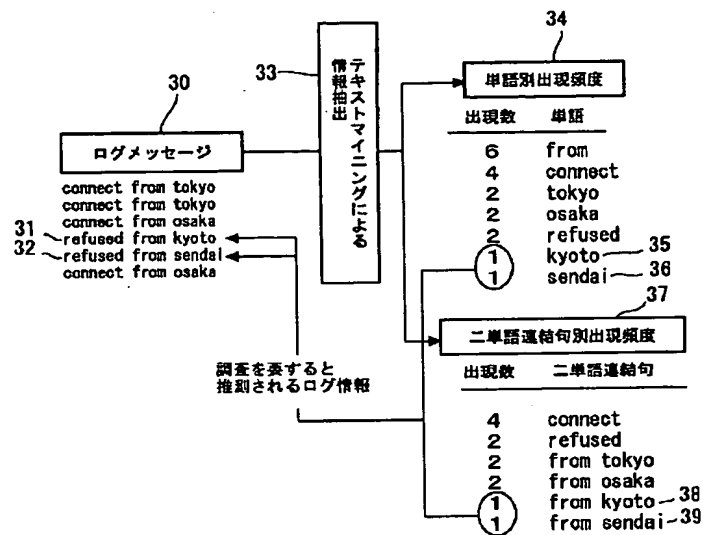
【図4】



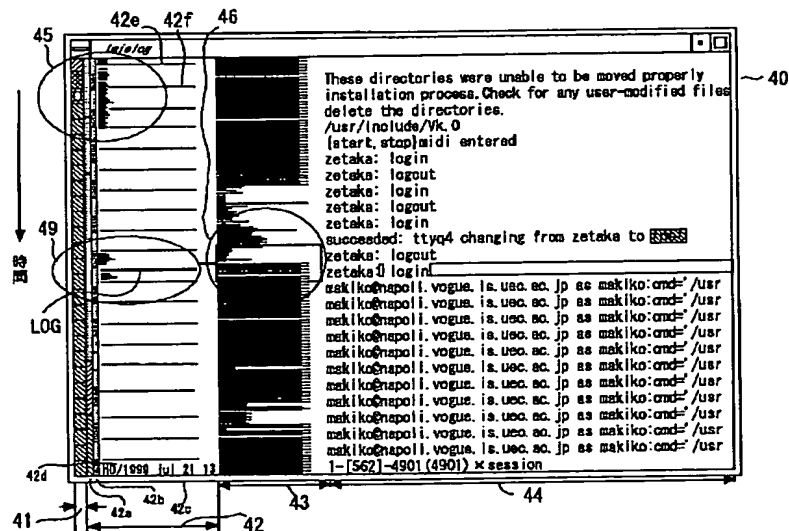
【図5】



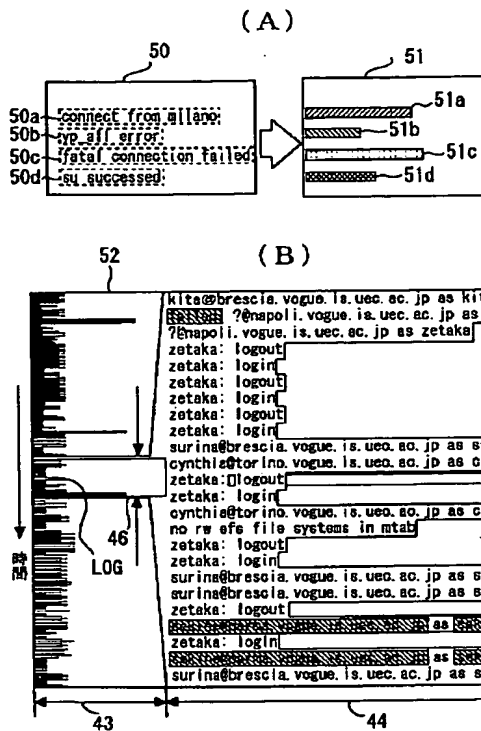
【図6】



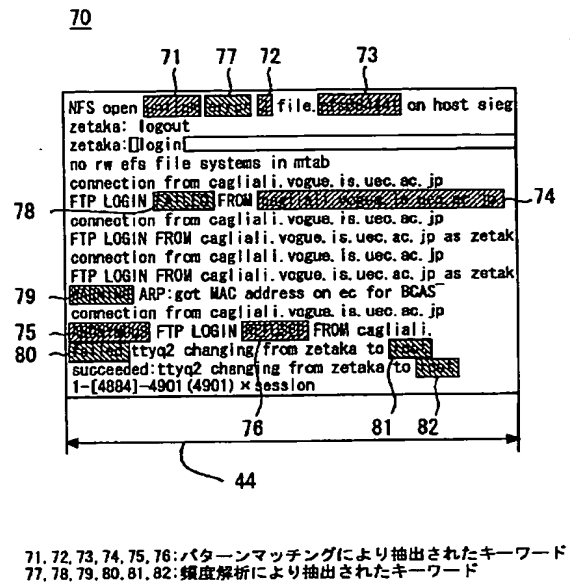
【図7】



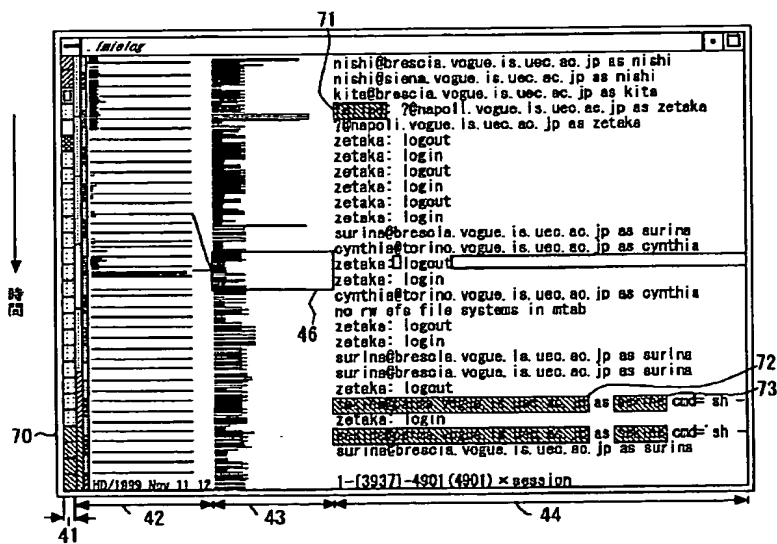
【図8】



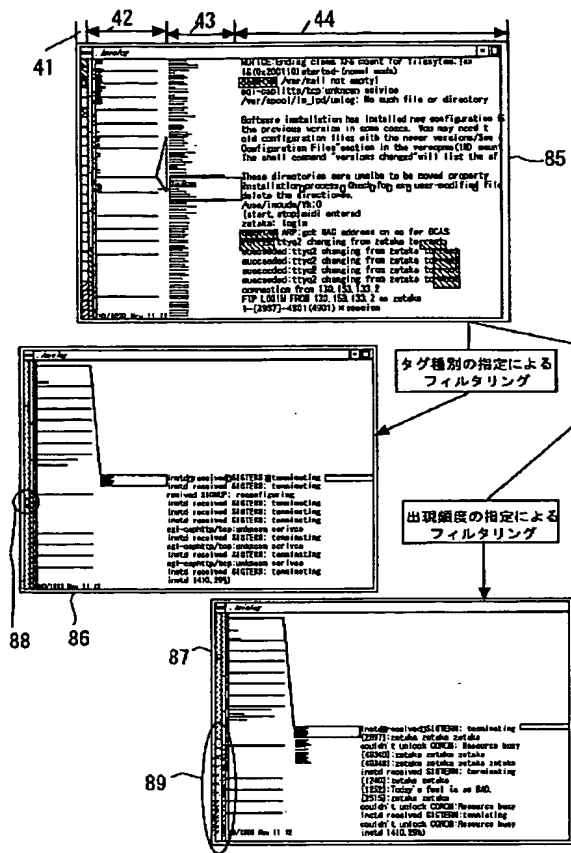
【図10】



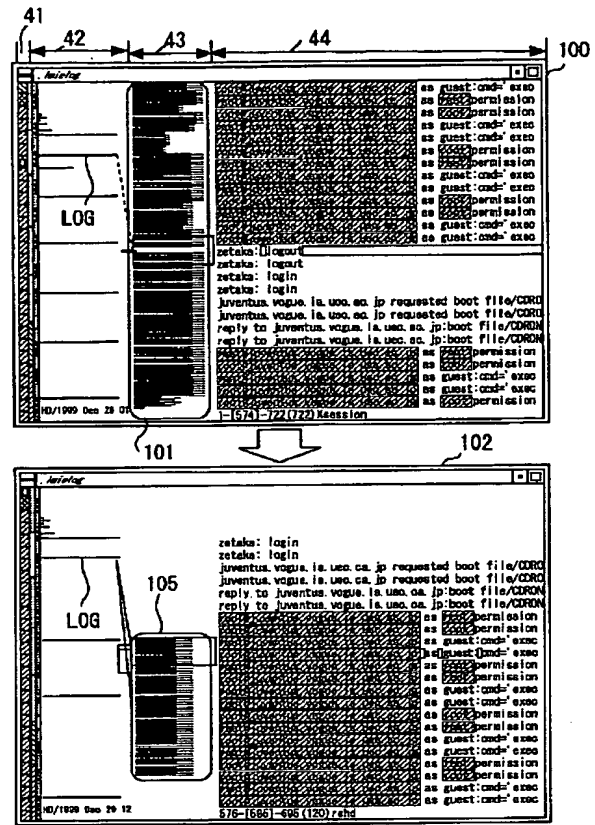
【図9】



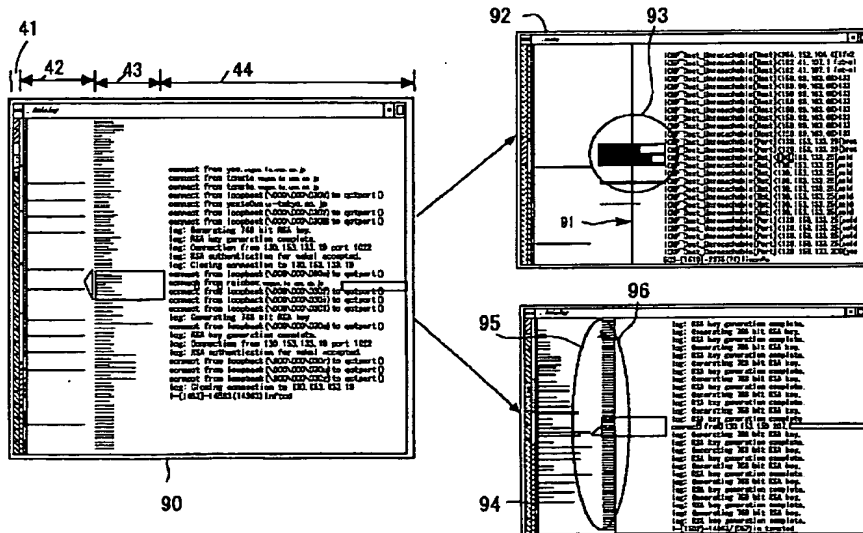
【図 11】



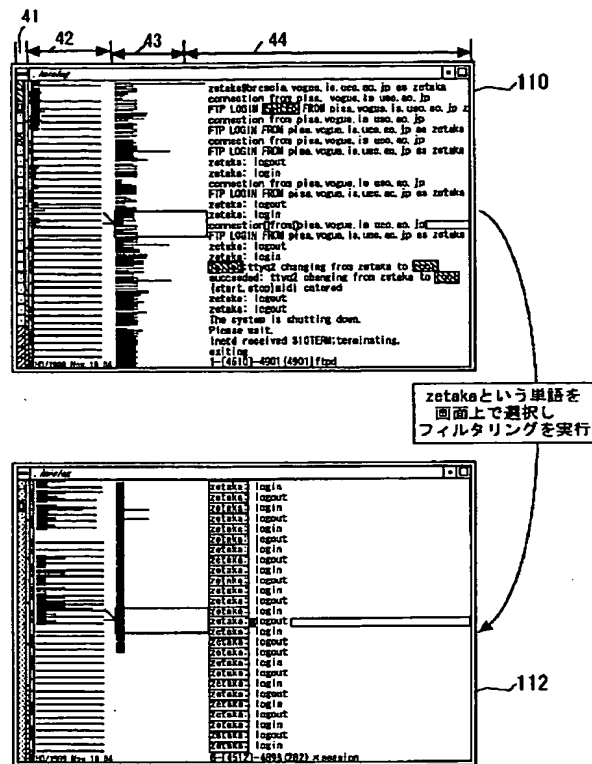
【図 13】



【図 12】



【図14】



フロントページの続き

(72)発明者 岡田 幹夫  
神奈川県横浜市鶴見区江ヶ崎町4番1号  
東京電力株式会社システム研究所内

Fターム(参考) 5B042 MA08 MA11 MA14 MC15 MC23  
MC28 MC35 MC40 NN04 NN09